Reducing dimensionality effect in importance sampling simulations

J.L. Sanz-González, S. Zazo and F. Álvarez-Vaquero

The dimensionality effect is avoided by the use of sufficient statistics in event probability estimators realised by importance sampling. If the system function is not a sufficient statistic, an approach is proposed to reduce the dimensionality effect in the estimators. Simulation results of false-alarm probability estimations, applied to radar detection, confirm a clear concordance with the theoretical results.

Introduction: As is well-known, the importance sampling (IS) technique [1–4] is a modified Monte-Carlo simulation applied to rare-event probability estimation, such as estimation of very low false-alarm probability in radar detection, or very low error-probability in communications.

Consider a random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ and its probability density function (PDF) $f_{\boldsymbol{X}}(\boldsymbol{x})$ defined in the R^n-space, being $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ a vector of real components. Define a statistic $t(\boldsymbol{x})$ and the critical region $t(\boldsymbol{x}) \geq t_0$ in the R^n-space; the probability $P_0 = \Pr\{t(\boldsymbol{X}) \geq t_0\} = E\{u(t(\boldsymbol{X}) - t_0)\}$, where $E\{\cdot\}$ means expectation with respect to $f_{\boldsymbol{X}}(\boldsymbol{x})$ and $u(\cdot)$ is the unit-step function. An alternative equation for this probability is $P_0 = E^*\{w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u(t(\boldsymbol{X}) - t_0)\}$, where $E^*\{\cdot\}$ means expectation with respect to $f^*_{\boldsymbol{X}}(\boldsymbol{x})$, known as the IS-PDF or the biasing PDF, and $w_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x})/f^*_{\boldsymbol{X}}(\boldsymbol{x})$ is the weighting function. An estimator of P_0 is the sample mean $\hat{P}_0$, i.e.

$$\hat{P}_0 = \frac{1}{N}\sum_{k=1}^{N} w_{\boldsymbol{X}}(\boldsymbol{x}^*_k) \cdot u\big(t(\boldsymbol{x}^*_k) - t_0\big) \tag{1}$$

being $\boldsymbol{x}^*_k$ and $k = 1, 2, \ldots, N$, independent sample vectors with PDF $f^*_{\boldsymbol{X}}(\boldsymbol{x})$.

If $f^*_{\boldsymbol{X}}(\boldsymbol{x}) \neq 0$ wherever $f_{\boldsymbol{X}}(\boldsymbol{x}) \neq 0$ in $t(\boldsymbol{x}) \geq t_0$, $\boldsymbol{x} \in R^n$, the estimator $\hat{P}_0$ is unbiased and consistent [1]. In fact, the variance $\sigma^2_{\hat{P}_0}$ of $\hat{P}_0$ is

$$\sigma^2_{\hat{P}_0} = E^*\Big\{\big(\hat{P}_0 - P_0\big)^2\Big\} = \frac{1}{N}\Big(E^*\Big\{\big[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u\big(t(\boldsymbol{X}) - t_0\big)\big]^2\Big\} - P_0^2\Big) \tag{2}$$

As is well-known [1], $f^*_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}) \cdot u\big(t(\boldsymbol{x}) - t_0\big)/P_0$ provides zero variance in (2) for any $N \geq 1$; however, it is not practical because P_0 is unknown (in fact, it has to be estimated). In the literature [1–4], some families of $f^*_{\boldsymbol{X}}(\boldsymbol{x})$ have been proposed for different estimation problems and the optimal solution is constrained to this family.

An important problem in many IS simulations is the dimensionality effect [2], meaning that $\sigma^2_{\hat{P}_0}$ increases as the dimension n of vector $\boldsymbol{x}$ increases, for a constant number N of simulation runs.

Reducing the dimensionality effect: If P_0 and N are fixed in (2), to minimise $E^*\{[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u(t(\boldsymbol{X}) - t_0)]^2\}$ is equivalent to minimising $\sigma^2_{\hat{P}_0}$.

Proposition 1. Define the weighting functions: $w_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x})/f^*_{\boldsymbol{X}}(\boldsymbol{x})$ *and* $w_T(t) = f_T(t)/f^*_T(t)$, *where* $t = t(\boldsymbol{x})$ *and the random variable* $T = t(\boldsymbol{X})$ *with* $f_T(\cdot)$ *and* $f^*_T(\cdot)$ *the corresponding induced PDF's; then, the following inequality holds:*

$$E^*\Big\{\big[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u(t(\boldsymbol{X}) - t_0)\big]^2\Big\} \geq E^*\Big\{\big[w_T(T) \cdot u(T - t_0)\big]^2\Big\} \tag{3}$$

Proof: Using conditional expectation, we can write

$$E^*\Big\{\big[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u\big(t(\boldsymbol{X}) - t_0\big)\big]^2\Big\} = E^*\Big\{E^*\Big\{\big[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u\big(t(\boldsymbol{X}) - t_0\big)\big]^2 | T\Big\}\Big\}$$

Considering that the variance is a non-negative number (or by Jensen's inequality), we can establish the following inequality:

$$E^*\Big\{\big[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u\big(t(\boldsymbol{X}) - t_0\big)\big]^2 | t\Big\} \geq \big[E^*\big\{w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u\big(t(\boldsymbol{X}) - t_0\big) | t\big\}\big]^2$$

Also, $E^*\big\{w_{\boldsymbol{X}}(\boldsymbol{X}) u\big(t(\boldsymbol{X}) - t_0\big) | t\big\} = \int_{R^n} w_{\boldsymbol{X}}(\boldsymbol{x}) u\big(t(\boldsymbol{x}) - t_0\big) f^*_{\boldsymbol{X}|T}(\boldsymbol{x}|t)\, d\boldsymbol{x}$ and $f^*_{\boldsymbol{X}|T}(\boldsymbol{x}|t) = f^*_{\boldsymbol{X}}(\boldsymbol{x}) f^*_{T|\boldsymbol{X}}(t|\boldsymbol{x})/f^*_T(t)$, $f^*_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = \delta\big(t - t(\boldsymbol{x})\big)$, $\delta(\cdot)$ being the Dirac delta function. Finally, after substituting the above, thus we have $E^*\{w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u(t(\boldsymbol{X}) - t_0)/t\} = w_T(t) \cdot u(t - t_0)$, and Proposition 1 has been proved.

Proposition 2: If $t(\boldsymbol{x})$ *is a sufficient statistic for the family of* $f^*_{\boldsymbol{X}}(\boldsymbol{x})$ *and* $f_{\boldsymbol{X}}(\boldsymbol{x})$, *then*

$$E^*\Big\{\big[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u\big(t(\boldsymbol{X}) - t_0\big)\big]^2\Big\} = E^*\Big\{\big[w_T(T) \cdot u(T - t_0)\big]^2\Big\} \tag{4}$$

Proof: Owing to the property of sufficiency: $f^*_{\boldsymbol{X}}(\boldsymbol{x}) = g(\boldsymbol{x}) \cdot h^*\big(t(\boldsymbol{x})\big)$ *and* $f_{\boldsymbol{X}}(\boldsymbol{x}) = g(\boldsymbol{x}) \cdot h(t(\boldsymbol{x}))$, *we have*

$$E^*\Big\{\big[w_{\boldsymbol{X}}(\boldsymbol{X}) \cdot u\big(t(\boldsymbol{X}) - t_0\big)\big]^2 | t\Big\} = \big[h(t)/h^*(t) \cdot u(t - t_0)\big]^2 \tag{5}$$

From the PDF definition (assuming $dt > 0$ and $dt \to 0$)

$$f_T(t)\, dt = \Pr\{t \leq t(\boldsymbol{X}) < t + dt\} = \int_{t \leq t(\boldsymbol{x}) < t + dt} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}$$
$$= h(t)\int_{t \leq t(\boldsymbol{x}) < t + dt} g(\boldsymbol{x})\, d\boldsymbol{x}$$

and

$$f^*_T(t)\, dt = h^*(t)\int_{t \leq t(\boldsymbol{x}) < t + dt} g(\boldsymbol{x})\, d\boldsymbol{x}$$

we can write $w_T(t) = f_T(t)/f^*_T(t) = h(t)/h^*(t)$ and its substitution in (5) finishes the proof.

Proposition 1 means that the variance of $\hat{P}_0$ realised by $w_{\boldsymbol{X}}(\boldsymbol{x})$ in (1) is greater than the variance of $\hat{P}_0$ realised by $w_T(t)$ in place of $w_{\boldsymbol{X}}(\boldsymbol{x})$, unless $t(\boldsymbol{x})$ is a sufficient statistic according to Proposition 2, where the dimensionality effect is cancelled.

Finally, if $t(\boldsymbol{x})$ is not a sufficient statistic, we should compute or estimate the weighting function $w_T(t)$, in order to be used in place of $w_{\boldsymbol{X}}(\boldsymbol{x})$ in (1) for reducing the dimensionality effect in the estimation of P_0. The computation of $w_T(t)$ should be performed through the computation of $f_T(\cdot)$ and $f^*_T(\cdot)$ over $t = t(\boldsymbol{x})$ from $f_{\boldsymbol{X}}(\boldsymbol{x})$ and $f^*_{\boldsymbol{X}}(\boldsymbol{x})$, respectively. This subject is now under research by the authors of this Letter.

Applications to radar detection: Consider the binary hypotheses H_0 (target absent) and H_1 (target present), defined as follows: $H_0{:}f_{\boldsymbol{X}}(\boldsymbol{x}|H_0) = \prod_{i=1}^{n} f_0(x_i)$ and $H_1{:}f_{\boldsymbol{X}}(\boldsymbol{x}|H_1) = \prod_{i=1}^{n} f_1(x_i)$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is the vector of n independent envelope samples (n is the number of pulses per antenna beam width), $f_0(x)$ is the PDF of the noise sample and $f_1(x)$ is the PDF of the target signal with noise. For the Gaussian noise, $f_0(x)$ is the Rayleigh distribution; for a non-fluctuating (NF) target model, $f_1(x)$ is the Rice distribution with S as the signal-to-noise ratio parameter; for the Swerling II (SW-II) target model [5], $f_1(x)$ is the Rayleigh distribution with parameter S_0 the signal-to-noise ratio mean. The optimum Neyman-Pearson detector has the well-known structure [5]

$$t(\boldsymbol{x}) = \sum_{i=1}^{n} a(x_i) \underset{H_0}{\overset{H_1}{\gtrless}} t_0 \tag{6}$$

where $a(x) \simeq x$ (linear) for a NF target model, and $a(x) = x^2$ (quadratic) for the SW-II target model. The false-alarm probability (P_{fa}) is defined by $P_{fa} = \Pr\{t(\boldsymbol{X}) \geq t_0/H_0\}$, hence, we can identify P_{fa} with P_0 and $f_{\boldsymbol{X}}(\boldsymbol{x}/H_0)$ with $f_{\boldsymbol{X}}(\boldsymbol{x})$ given above. According to Orsak [3], a good IS biasing PDF is $f^*_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}|H_1)$ with an adequate signal-to-noise ratio. Since the linear statistic is a sufficient statistic for detecting a NF target model, the adequate IS biasing PDF should correspond to the NF case, in order to cancel the dimensionality effect; on the other hand, quadratic statistic is a sufficient statistic for the SW-II target model, and the adequate IS biasing PDF should correspond to the SW-II case for cancelling the dimensionality effect.

Simulation results: Now, we consider the linear and quadratic statistics as detector statistics in (6), and estimate the false-alarm probability by using IS PDF's corresponding to the NF and SW-II target models. The results are shown in Figs. 1 and 2, where the parameter values are: number of integrated pulses, $n = 8$; number of simulation runs, $N = 1000$; IS-biasing parameters, $S = S_0 = 3$. In Fig. 1, we show P_{fa}-estimations against threshold t_0 for both detectors, using both biasing PDF's: NF (NF: dotted lines) and SW-II (SW-II: solid lines); in addition, we show theoretical curves (discontinuous red lines) of

P_{fa}-values (computed from integral equations). In Fig. 2, we show the corresponding relative-error estimation ($\hat{\varepsilon}$) of P_{fa}-estimations:

$$\hat{\varepsilon} = \sqrt{\hat{\sigma}^2_{\hat{P}_{fa}}} / \hat{P}_{fa}$$

for the four possibilities, where $\hat{\sigma}^2_{\hat{P}_{fa}}$ is a variance estimator of $\hat{P}_{fa}$. From Figs. 1 and 2, it can be seen that for the quadratic detector, $\hat{P}_{fa}$-values obtained by SW-II as IS-PDF are very close to the exact P_{fa}-values, and better than those obtained by the NF procedure. On the contrary, for the linear detector, the $\hat{P}_{fa}$ provided by a NF as IS-PDF is better than that provided by the SW-II procedure. Similar results are obtained for $n = 16$, 32 or 64.

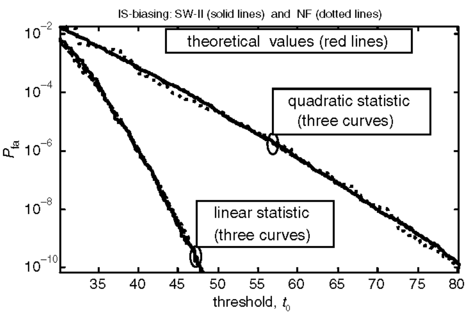

Fig. 1 *False-alarm probability (P_{fa}) against threshold t_0 for linear and quadratic detectors: theoretical values and estimations by two different IS-PDF's (SW-II and NF)*

Parameters: $n = 8$, $S = S_0 = 3$ and $N = 1000$

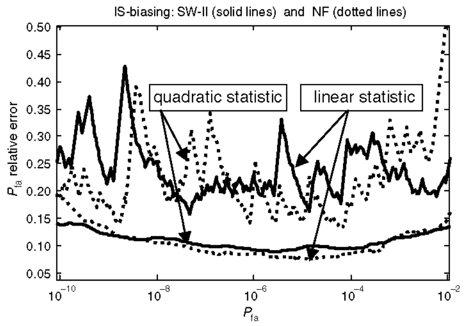

Fig. 2 *P_{fa}-relative error ($\hat{\varepsilon}$) against P_{fa} in same conditions of Fig. 1*

Conclusion: It has been shown that if the system statistic is sufficient, the dimensionality effect is circumvented in rare-event probability estimation by the IS technique. Applications to false-alarm probability estimations in radar detectors have been given, and the simulation results are in agreement with the theoretical ones.

Acknowledgements: This work was supported by the Ministerio de Ciencia e Innovación (grant TEC2010-21217-C02-02/TCM), and the 'Universidad Politécnica de Madrid.

26 April 2013

doi: 10.1049/el.2013.1451

One or more of the Figures in this Letter are available in colour online.

J.L. Sanz-González, S. Zazo and F. Álvarez-Vaquero (*Dpto. SSR, Universidad Politécnica de Madrid, ETSI de Telecomunicación-UPM, Ciudad Universitaria, Madrid 28040, Spain*)

E-mail: jlsanz@gcs.ssr.upm.es

References

1 Chen, J.C., Lu, D., Sadowsky, J.S., and Yao, K.: 'On importance sampling in digital communications – Part I: Fundamentals', *IEEE J. Sel. Areas Commun.*, 1993, **11**, (3), pp. 289–299

2 Smith, P.J., Shafi, M., and Gao, H.: 'Quick simulation: a review of importance sampling techniques in communications systems', *IEEE J. Sel. Areas Commun.*, 1997, **15**, (4), pp. 597–613

3 Orsak, G.C.: 'A note on estimating false alarm rates via importance sampling', *IEEE Trans. Commun.*, 1993, **41**, (9), pp. 1275–1277

4 Vicen-Bueno, R., Jarabo-Amores, M.P., Rosa-Zurera, M., Sanz-González, J.L., and Maldonado-Bascón, S.: 'Importance sampling for objective function estimations in neural detector training driven by genetic algorithms', *Neural Process. Lett.*, 2010, **32**, (3), pp. 249–268

5 Peebles, P.Z.: 'Radar principles' (John Wiley & Sons, Inc., New York, 1998)